

# Functional mechanisms involved in the internal inhibition of taboo words

Els Severens,<sup>1,2,\*</sup> Simone Kühn,<sup>1,2,\*</sup> Robert J. Hartsuiker,<sup>1,2</sup> and Marcel Brass<sup>1,2</sup>

<sup>1</sup>Department of Experimental Psychology and <sup>2</sup>Ghent Institute for Functional and Metabolic Imaging, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

**The present study used functional magnetic resonance imaging to investigate brain processes associated with the inhibition of socially undesirable speech. It is tested whether the inhibition of undesirable speech is solely related to brain areas associated with classical stop signal tasks or rather also involves brain areas involved in endogenous self-control. During the experiment, subjects had to do a SLIP task, which was designed to elicit taboo or neutral spoonerisms. Here we show that the internal inhibition of taboo words activates the right inferior frontal gyrus, an area that has previously been associated with externally triggered inhibition. This finding strongly suggests that external social rules become internalized and act as a stop-signal.**

**Keywords:** fMRI; inhibition; speech; rIFG; taboo words

## INTRODUCTION

When acting in public, most people try to behave in a socially desirable way to avoid embarrassing situations. Being in disagreement with the social rules can be very embarrassing. For example, if someone produces a slip of the tongue, which accidentally results in a taboo word, this is considered very embarrassing for the speaker. As children, we learn appropriate social behavior from our environment. For instance, parents teach children that it is inappropriate to curse, especially in public. In these cases, parents provide an external signal indicating to the child not to talk dirty. As grown-ups, we usually know what behavior is inappropriate and have learned to inhibit these actions. But what exactly are the mechanisms that prevent adults from behaving socially inappropriate? Does the inhibition of social undesirable behavior involve the same functional and neural mechanisms that are involved in the inhibition of behavior that is neutral with respect to social desirability or does it involve an additional act of deliberation?

While the question of how socialization influences our behavior has already been raised a while ago in psychology (Freud, 1961) and has been debated ever since (e.g. Pittet, 2004). Modern brain-imaging techniques can provide a new perspective on this question. Functional brain imaging allows investigating cognitive operations that might not become manifest in overt behavior. Recent brain-imaging research demonstrated that stopping of manual responses in stop-signal paradigms relies on a network of brain regions

of which the most crucial component is located in the right inferior frontal gyrus (rIFG; e.g. Aron *et al.*, 2007). Recently, Xue *et al.* (2008) found that the rIFG is also active with the inhibition of verbal responses. On the other hand, inhibition that is not guided by an external cue but rather internally guided (endogenous self-control), has been demonstrated to involve the medial prefrontal cortex, more specifically the dorsal fronto-median cortex (dFMC; Brass and Haggard, 2007, 2008; Kühn *et al.*, 2009). Generally, definitions of self-control do not distinguish between these two forms of inhibition but rather define self-control more generally as 'the overriding or inhibiting of automatic, habitual, or innate behaviors, urges, emotions, or desires that would otherwise interfere with goal directed behavior' (Muraven *et al.*, 2006). We think, however, that it is very crucial to distinguish endogenous from externally guided inhibition, both conceptually and neuroanatomically. Externally guided inhibition is not a result of deliberation but is rather triggered by the environment. By contrast, endogenous self-control is related to a deliberate decision whether to act or not and therefore activates brain areas that are involved in decision making (Brass and Haggard, 2007, 2008; Kühn *et al.*, 2009). While the concept of endogenous inhibition was developed in the motor domain, a recent study of S. Kühn *et al.* (unpublished data) demonstrated that this distinction is also found for the inhibition of emotions.

If the inhibition of taboo words simply relies on the classical inhibition network, one would expect the rIFG to be involved solely. However, if it is based on a process of endogenous self-control one would expect additional brain activation in the dFMC.

The present study investigated brain processes associated with the inhibition of taboo words. We used the SLIP task,

Received 19 October 2010; Accepted 10 April 2011

This work was supported by the Research Foundation Flanders (FWO-Vlaanderen) to E.S. and S.K. and by the EUROVETO project (09-ECRP-020) of the European Science Foundation.

Correspondence should be addressed to Els Severens, Department of Experimental Psychology and Ghent Institute for Functional and Metabolic Imaging, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: els.severens@ugent.be

\*These authors contributed equally to this work.

which is well-known to elicit speech errors in the language production literature (Baars *et al.*, 1975; Hartsuiker *et al.*, 2005). SLIP tasks have been used earlier to study inhibition of taboo words (Motley *et al.*, 1981). In this task, participants have to read word pairs and occasionally have to pronounce the word pair after reading. The phonological composition of the preceding word pairs elicits spoonerisms, namely exchanges of the first phonemes of the two words (e.g. *mad dash > dad mash*). The resulting spoonerisms in the present study were either taboo or neutral words as established in a pretest. In a previous ERP study (Severens *et al.*, in press), also comparing taboo and neutral spoonerisms, we demonstrated that word pairs that potentially resulted in taboo spoonerisms (taboo-eliciting pairs) elicited conflict ‘even though no overt errors were committed’. We concluded that there is stronger inhibition of taboo words than of neutral words. However, the ERP study could not answer the question whether inhibition of taboo words activates the same brain circuits as the inhibition of neutral words to a higher degree (rIFG) or whether taboo words involve endogenous self-control, as reflected in activation of the dFMC. The aim of the current event-related functional magnetic resonance imaging (fMRI) study was to address this question using virtually the same paradigm that has already been used in the EEG study.

## METHODS

### Subjects

We recruited 17 healthy native Dutch speakers (5 males; age: mean = 22.2, ranging from 19 to 27) from whom we obtained written consent prior to the scanning session. All subjects had normal or corrected-to-normal vision. No subject had a history of neurological, major medical or psychiatric disorder. All subjects were right handed as assessed by the Edinburgh handedness questionnaire (Oldfield, 1971).

### Behavioral task

One hundred and twenty-two word pairs were constructed. When exchanging the first phonemes of these target word pairs, half of them formed a taboo pair (e.g. *katten nut* → *natte k\*t*; *cats sense* → *wet c\*nt*), and the other half created a neutral pair [e.g. *katten nok > natte kok (wet chef)*]. Taboo and neutral target pairs started with the same phonemes; furthermore, they were matched on syntactic structure and frequency. The taboo and neutral word pairs were judged by 46 participants on a 7-point ‘tabooness’ scale (1 = neutral and 7 = taboo; taboo: 5.71; neutral: 2.12). For both taboo and neutral pairs, three biasing pairs were constructed; these word pairs started with the same phonemes as the taboo and neutral pairs. Thus the phonemes were exchanged in the target pairs (participants only saw the target pairs).

We presented word pairs for 800 ms in red on a black background, which had to be read silently (Figure 1). Each trial consisted of the presentation of a fixation cross for 500 ms and 2–7 word pairs presented subsequently with an

interstimulus interval of 1000 ms and a variable intertrial interval of 0, 500, 1000 or 1500 ms. An exclamation mark (!, for 2500 ms) following the target word pair indicated that participants had to vocalize the preceding word pair as fast as possible. Responding was monitored via a camera.

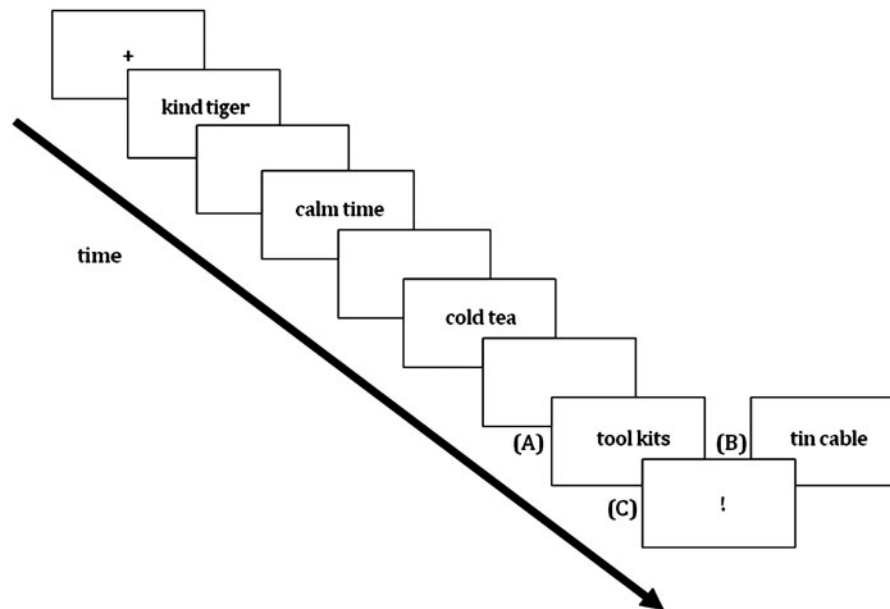
The experiment comprised 63 taboo-eliciting target pairs and 63 neutral target pairs of which 13 trials per condition were followed by the exclamation mark. These trials were included to make sure the participants prepare the word pairs to be pronounced and to make them more prone to internally generate the spoonerism. In addition, we included 63 nullevnts. In total, the experiment consisted of 3 runs and lasted ~30 min.

### Scanning procedure

Images were collected with a 3T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany) using an 8-channel radiofrequency head coil. First, high-resolution anatomical images were acquired using a T1-weighted 3D MPRAGE sequence (TR = 1550 ms, TE = 2.39 ms, TI = 900 ms, acquisition matrix = 256 × 256 × 176, sagittal FOV = 220 mm, flip angle = 9°, voxel size = 0.9 × 0.9 × 0.9 mm<sup>3</sup>). Whole-brain functional images were collected using a T2\*-weighted EPI sequence sensitive to BOLD contrast (TR = 2000 ms, TE = 35 ms, image matrix = 64 × 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 17%, voxel size 3.5 × 3.5 × 3 mm<sup>3</sup>, 30 axial slices). Approximately, 250 image volumes aligned to AC–PC were acquired per run.

### fMRI data pre-processing and GLM analysis

The fMRI data were analyzed with statistical parametric mapping using the SPM5 software (Wellcome Department of Cognitive Neurology, London, UK). The first four volumes of all EPI series were excluded from the analysis to allow the magnetization to approach a dynamic equilibrium. Data processing started with slice time correction and realignment of the EPI data sets. A mean image for all EPI volumes was created, to which individual volumes were spatially realigned by rigid body transformations. The high-resolution structural image was co-registered with the mean image of the EPI series. Then the structural image was normalized to the Montreal Neurological Institute (MNI) template, and the normalization parameters were applied to the EPI images to ensure an anatomically informed normalization. During normalization, the anatomy image volumes were regridded to 1 × 1 × 1 mm<sup>3</sup>. A commonly applied spatial filter of 8 mm FWHM (full-width at half maximum) was used. Low-frequency drifts in the time domain were removed by modeling the time series for each voxel by a set of discrete cosine functions to which a cut-off of 128 s was applied. The subject-level statistical analyses were performed using the general linear model (GLM). We modeled the taboo-eliciting pairs, neutral pairs, taboo-eliciting pairs with subsequent exclamation, neutral pairs with subsequent



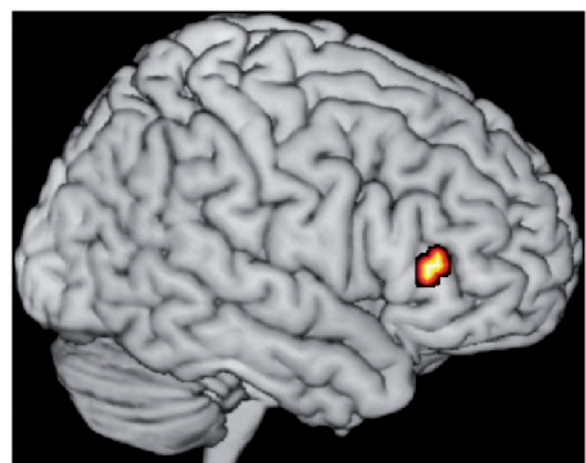
**Fig. 1** Schematic drawing of the SLIP task. (A) taboo-eliciting trial, (B) neutral trial and (C) pronunciation cue.

exclamation and null-events separately. Vectors containing the event onsets were convolved with the canonical hemodynamic response function (HRF) to form the main regressors in the design matrix (the regression model). The vectors were also convolved with the temporal derivatives and the resulting vectors were entered into the model. The statistical parameter estimates were computed separately for each voxel for all columns in the design matrix. Contrast images were constructed from each individual to compare the relevant parameter estimates for the regressors containing the canonical HRF. Resulting statistical values were thresholded at  $P < 0.001$  (uncorrected) with a volume greater than  $175 \text{ mm}^3$  (5 adjacent voxels). They were overlaid onto a normalized T1-weighted MNI single subject template (colin27).

Percent signal change analyses were carried out in the dmFC using MARSBAR. The exact area was BA9 at MNI coordinates  $-2 \ 41 \ 34$ . This is the peak coordinate that has been found in the study of Brass and Haggard (2007). The region of interest (ROI) was defined as a sphere with a radius of 6 mm around the peak coordinate. For each subject and condition separately, the mean percent signal change over a time window of 4–6 s after stimulus onset was calculated and compared by means of paired  $t$ -tests.

## RESULTS

In order to explore the neural correlates of the inhibition of taboo words, we focused on the brain activity that was increased in taboo-eliciting trials compared to neutral trials (only trials without an exclamation mark). Speech artifacts contaminated the trials with exclamation marks. Inhibition of socially undesirable stimuli more strongly activated the rIFG (IFG, BA 45, MNI coordinates: 56, 35, 11, Figure 2).



**Fig. 2** Main contrast of taboo-eliciting vs neutral conditions. Activation map averaged over 17 subjects ( $P < 0.001$ ,  $k = 5$  voxel) mapped onto a T1 image in MNI space. Displayed is an activity in right inferior frontal gyrus (IFG, BA 45).

Previous studies showed that the pars triangularis (BA 45), which is part of the rIFG, is related to externally triggered inhibition (Aron *et al.*, 2003, 2007).

To examine whether there is indeed no additional activation of the dmFC in the taboo condition, an ROI analysis was carried out in this area (BA 9, MNI coordinates:  $-2 \ 41 \ 34$ ). We found no difference in the activation of the dmFC between the taboo and the neutral condition [ $t(16) = -1.33$ ,  $P > 0.20$ ].

Finally, to show that no other brain areas related to self-control are active, we lowered the threshold to  $P < 0.005$ . This yielded an additional activation focus in the left IFG, but there was no further activation.

## DISCUSSION

The present study used a SLIP task to investigate the functional and neural mechanisms involved in the inhibition of taboo words. We argued that if the inhibition of socially undesirable stimuli is based on general inhibition mechanisms, the inhibition of taboo words should exclusively activate brain areas that are also found for the inhibition of socially neutral stimuli. In line with this hypothesis, the inhibition of taboo words led to increased activation of the rIFG compared to inhibition of neutral words. No activation, however, was found in the medial prefrontal cortex as would have been predicted if the inhibition of taboo words would involve an act of endogenous self-control.

The rIFG plays a crucial role for the regulation of behavior. In particular, it has been associated with motor control (e.g. Aron *et al.*, 2003), control of risky behavior (e.g. Engemann and Tamir, 2009), temporal discounting (e.g. Wittman *et al.*, 2007) and emotion regulation (Kim and Hamann, 2007). Importantly, the rIFG has been proposed to implement an active mechanism of neural inhibition. Accordingly, several studies have associated the rIFG with inhibition of manual responses (Aron *et al.*, 2003, 2007; Chambers *et al.*, 2006) and recently also with the inhibition of verbal responses (Xue *et al.*, 2008). Therefore, the present data strongly suggest that the mechanisms involved in the inhibition of taboo words are similar to the mechanisms involved in the inhibition of socially neutral stimuli.

In most studies on response inhibition, manual responses had to be inhibited upon detection of a stop signal (Aron *et al.*, 2007). Recently, Xue *et al.* (2008) found that the inhibition of verbal responses activates the same brain areas as the inhibition of manual responses, namely, the rIFG. In these studies, an external sign always signaled both verbal and manual stops. The present study investigated a more natural way of inhibiting verbal responses; no overt signal cued the participant to inhibit a verbal response. Importantly, in principle, participants did not need to inhibit taboo spoonerisms more than neutral spoonerisms because neither were correct responses. Nevertheless, rIFG activation demonstrated a difference between these two conditions, presumably because taboo words are socially unacceptable. People want to act in a socially desirable way and do not want to use taboo words in public. Hence inhibition of taboo words is stronger. In a previous ERP study, the data also supported this claim (Severens *et al.*, in press). Importantly, these findings are in line with theories of speech monitoring which generally assume that speakers can inspect inner speech and inhibit speech plans that are incorrect or inappropriate before these become overt (Motley *et al.*, 1981; Levelt, 1989; Postma, 2000; Hartsuiker and Kolk, 2001).

When growing up there are unquestionably external signals that tell the child not to use taboo words, especially, from the parents. As an adult, there is no one who tells you not to utter taboo words, but it is socially undesirable to do

so. Recently, Anderson and colleagues (2001, 2004) showed that voluntary suppression of words can become unintentional. Similarly, our data suggest that external stopping can become independent from explicit deliberation. How might such an internal stopping mechanism operate? It is possible that the contingent coupling of a word with an external stop-command leads to an association of the word and the stop-command so that thinking the taboo word automatically activates stopping. A similar mechanism was recently demonstrated by Verbruggen and Logan (2008).

Our findings are also in accordance with research in the clinical population. People with Tourette Syndrome (TS) have motor or vocal tics, which they are unable to inhibit. Interestingly, these tics are often seen as socially undesirable. It has been suggested that TS patients have deficits in cognitive control (Watkins *et al.*, 2005). On the neuroanatomical level, TS has been related to reduced white matter in the rIFC (Müller-Vahl *et al.*, 2009). Furthermore, when children with TS are performing cognitive control tasks there is more activity in the rIFG, yet behaviorally they perform on the same level as their matched controls do (Baym *et al.*, 2008). Baym *et al.* interpreted these results as suggesting that TS patients overcompensate to inhibit certain responses. These data are consistent with our findings that the rIFG is related to the inhibition of socially undesirable behavior.

To conclude, the present study showed that the inhibition of the taboo word errors is stronger than the inhibition of the neutral errors. Furthermore, it showed that the mechanisms involved in the inhibition of taboo words are similar mechanisms involved that are involved in the classical stop-signal task. This shows that external social rules become internalized and act as a stop signal.

## Conflict of Interest

None declared.

## REFERENCES

- Anderson, M.C., Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410(6826), 366–9.
- Anderson, M.C., Ochsner, K.N., Kuhl, B., et al. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, 303(5655), 232–5.
- Aron, A.R., Behrens, T.E., Smith, S., Frank, M.J., Poldrack, R.A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience*, 27(14), 3743–52.
- Aron, A.R., Fletcher, P.C., Bullmore, E.T., Sahakian, B.J., Robbins, T.W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, 6(12), 115–6.
- Baars, B.J., Motley, J.T., MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–91.
- Baym, C.L., Corbett, B.A., Wright, S.B., Bunge, S.A. (2008). Neural correlates of tic severity and cognitive control in children with Tourette syndrome. *Brain*, 131, 165–79.
- Brass, M., Haggard, P. (2007). To do or not to do: The neural signature of self-control. *Journal of Neuroscience*, 27(34), 9141–5.
- Brass, M., Haggard, P. (2008). The What, When, Whether model of intentional action. *Neuroscientist*, 14(4), 319–25.



- Chambers, C.D., Bellgrove, M.A., Stokes, M.G., et al. (2006). Executive “brake failure” following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience*, 18(3), 444–55.
- Engelmann, J.B., Tamir, D. (2009). Individual differences in risk preference predict neural responses during financial decision-making. *Brain Research*, 1290, 28–51.
- Freud, S. (1961). In Strachey, J., editor. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 19, London: Hogarth Press.
- Hartsuiker, R.J., Corley, M., Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn’t fly: related reply to Baars et al. (1975). *Journal of Memory and Language*, 52(1), 58–70.
- Hartsuiker, R.J., Kolk, H.H.J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology*, 42(2), 113–57.
- Kim, S.H., Hamann, S. (2007). Neural correlates of positive and negative emotion regulation. *Journal of Cognitive Neuroscience*, 19(5), 776–98.
- Kühn, S., Haggard, P., Brass, M. (2009). Intentional inhibition: how the “veto-area” exerts control. *Human Brain Mapping*, 30(9), 2834–43.
- Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Motley, M.T., Camden, C.T., Baars, B.J. (1981). Toward verifying the assumptions of laboratory-induced slips of the tongue—the output-error and editing issues. *Human Communication Research*, 8(1), 3–15.
- Müller-Vahl, K.R., Kaufmann, J., Grosskreutz, J., Dengler, R., Emrich, H.M., Peschel, T. (2009). Prefrontal and anterior cingulate cortex abnormalities in Tourette Syndrome: evidence from voxel-based morphometry and magnetization transfer imaging. *BMC Neuroscience*, 10, 47.
- Muraven, M., Shmueli, D., Burkley, E. (2006). Conserving self-control strength. *Journal of Personality and Social Psychology*, 91(3), 524–37.
- Oldfield, R.C. (1971). Assessment and analysis of handedness—Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Pittet, D. (2004). The lowbury lecture: behaviour in infection control. *Journal of Hospital Infection*, 58, 1–13.
- Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77(2), 97–131.
- Severens, E., Janssens, I., Kühn, S., Brass, M., Hartsuiker, R.J. (in press). When the brain tames the tongue: Covert editing of inappropriate language. *Psychophysiology*.
- Verbruggen, F., Logan, G.D. (2008). Automatic and controlled response inhibition: associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology-General*, 137(4), 649–72.
- Watkins, L.H., Sahakian, B.J., Robertson, M.M., et al. (2005). Executive function in Tourette’s syndrome and obsessive-compulsive disorder. *Psychological Medicine*, 35(4), 571–82.
- Wittmann, M., Leland, D.S., Paulus, M.P. (2007). Time and decision making: differential contribution of the posterior insular cortex and the striatum during delay discounting task. *Experimental Brain Research*, 179(4), 643–53.
- Xue, G., Aron, A.R., Poldrack, R.A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, 18(8), 1923–32.